Educación y Ciencia

EDITORIAL

UADY FACULTAD DE EDUCACIÓN

http://www.educacionyciencia.org

Sobre la Técnica de Puntajes de Propensión (Propensity Score Matching) y sus usos en la investigación en Educación

Reviewing of Propensity Score Matching and its Uses in Educational Research

Claudia Patricia Ovalle Ramírez

Universidad de Antioquia, Colombia.

Información adicional sobre este manuscrito escribir a:

Claudia Ovalle Ramírez, claudia.ovalle@udea.edu.co

Cómo citar este artículo:

Ovalle Ramírez, C. (2015). Sobre la técnica de Puntajes de Propensión (Propensity Score Matching) y sus usos en la investigación en Educación. *Educación y Ciencia*, *4*(43), 80-89.

Resumen

La técnica de Propensity Score Matching no es nueva, pero su uso no ha sido extenso en la investigación en Educación. En el presente artículo se argumenta que la utilización de esta técnica es conveniente en investigación en educación porque permite solucionar dificultades como el sesgo de selección y la imposibilidad de llevar a cabo experimentos por motivos éticos u otros. En ese sentido desarrollar diseños cuasi-experimentales ayudados con técnicas estadísticas como Propensity Score Matching (PSM) puede ser una alternativa viable que arroja resultados más precisos (al nivel de los hallados en situaciones experimentales) sobre efectos de interés.

Palabras clave: Propensity Score Matching, investigación, diseño cuasi-experimental, educación, causalidad

Abstract

Propensity Score Matching is not a recent statistical technique yet. It has not been used extensively in educational research. In the following paper it is stated that the technique would be worth of use as researchers may not always have the resources or ideal conditions in order to carry out experiments because of ethical and resource constraints. Quasi-experimental designs may be an alternative approach. By using PSM (Propensity Score Matching) it is possible to obtain more accurate results.

Keywords: Propensity Score Matching, research, quasi-experimental design, education, causality

Introducción

El Emparejamiento por Puntajes de Propensión o Propensity Score Matching (PSM) tiene el propósito de fortalecer los argumentos sobre la causalidad de las relaciones entre variables en estudios cuasi-experimentales. Esto debido a que el PSM reduce el sesgo de selección, es decir, el sesgo que se genera porque no se pueden asignar de forma aleatoria los individuos a una intervención o a un grupo control, para balancear todas aquellas características que pueden afectar la estimación del efecto.

Debido a que es un método *ex post-fact*o, el PSM empareja los casos que recibieron una intervención (como si fueran parte de un grupo tratamiento experimental) con uno o más casos que no reciben dicha intervención. En este sentido, la diferencia con un experimento radica en que el grupo control puede ser “creado” con base a la información disponible (por ejemplo, un grupo control creado a partir de datos censales de personas que no son participantes). Incluso en diseños experimentales donde se contempla un grupo experimental y uno control aleatoriamente asignados, el PSM puede servir para balancear aún más las diferencias entre estos grupos permitiendo mejores estimaciones.

El emparejamiento de los casos se hace por medio de puntajes de propensión. Estos puntajes son probabilidades calculadas con base a características observables de los casos que reducen las diferencias entre estos con independencia de que hayan recibido o no el tratamiento, al igual que ocurre con la asignación aleatoria en los experimentos controlados. En palabras de Vinha (2006): “Los estimadores no paramétricos de emparejamiento (puntajes de propensión) son empleados frecuentemente en estudios de evaluación. La idea general de esta metodología consiste en determinar el impacto del tratamiento en los casos tratados usando información de observaciones que han recibido tratamiento y de observaciones similares sin tratamiento para construir un contra-factual sin tratamiento” (p.21).

Por su parte Heinrich, Maffioli y Vasquez (2010) indican que PSM “usa información de grupo de unidades que no participan en la intervención para identificar que le habría pasado a las unidades participantes en la ausencia de intervención. Por medio de la comparación de cómo los resultados difieren entre los participantes en comparación con no-participantes que son equivalentes en sus características observables, es posible estimar los efectos de la intervención promediando las diferencias entre participantes y sus casos de comparación emparejados” (p.6).

El PSM es una técnica cuyos resultados son bastante aproximados a los obtenidos en las situaciones experimentales reales, en términos del cálculo de los impactos (de una variable independiente sobre una dependiente) y de sus errores estándar, pero evitando las dificultades (éticas y de recursos) de una situación experimental real. Estos beneficios se han comprobado en estudios que comparan resultados obtenidos con medios experimentales con los de grupos de comparación no obtenidos aleatoriamente (i.e., estudios observacionales).

Por ejemplo, en el artículo de Dehejia y Wahba (1999) se reporta la evaluación del efecto de un programa de entrenamiento laboral sobre el salario de los participantes haciendo comparaciones entre diferentes grupos: grupo tratamiento y grupo control (ambos obtenidos de forma aleatoria), y de estos con muestras de tipo observacional de personas que no son participantes del programa. Se compara inicialmente el grupo tratamiento con el grupo control, ambos compuestos por participantes del programa -pero diferenciados en la época de reclutamiento-. Se encuentra que existe un efecto en términos de una ganancia salarial de los sujetos del grupo tratamiento equivalente a US $1,794 en comparación con el grupo control tras su participación en el programa. Este

efecto fue a su vez comparado con el efecto estimado en muestras de personas que no son participantes del programa, y cuya información socio-demográfica y de ganancias salariales se reporta en encuestas nacionales. Para el subgrupo de participantes de la encuesta con sexo masculino el efecto estimado (si hubieran recibido el programa) llega a US $1,473, mientras que este llega a los US $1,774 dólares para la muestra total de encuestados. Los efectos hallados en las muestras observacionales (encuesta) se aproximan bastante a los del grupo experimental, por lo cual se puede concluir que con los datos apropiados (características socio-demográficas, variables de respuesta como el salario, entre otros) los efectos pueden ser estimados en diseños observacionales como lo haría un diseño experimental.

Usos del PSM en educación

Estimaciones de los efectos como las del ejemplo anterior se logran por medio de técnicas como PSM. Esta técnica puede ser ampliamente usada en la investigación en educación para evitar las dificultades que se presentan con el sesgo de selección y con el desarrollo de paradigmas de experimentación en este tipo de investigación. En el campo de la educación sería éticamente incorrecto asignar niños a una institución educativa de baja calidad para poder analizar su efecto en variables como el desempeño del estudiante, por lo cual existe un sesgo antes de realizar cualquier análisis. Este sesgo consiste en que los estudiantes pueden presentan una serie de características previas al ingreso a la institución las cuales pueden afectar sus resultados académicos como el nivel educativo de los padres, la zona de residencia y las condiciones económicas. Estas son covariables que afectan el rendimiento académico (i.e., el efecto del barrio, del capital cultural, etc.) pero que no son los efectos que interesan al investigador y que pueden sesgar los resultados que éste obtiene ya que se puede confundir su efecto con el de las variables de tratamiento (i.e., calidad del colegio).

En la asignación aleatoria propia del paradigma experimental las características de los individuos son balanceadas por medio de la asignación al azar de los casos entre el grupo experimental y el grupo control. Para evitar los sesgos, estos grupos no deben diferir en otras características a excepción de las producidas por la intervención. Una alternativa a la imposibilidad de asignar sujetos a un tratamiento (como escuela de baja calidad vs alta calidad; colegio de jornada completa vs. jornada parcial; institución educativa pública vs. privada) consiste en establecer que las distribuciones de características (covariables) de los individuos son equivalentes, para luego estimar el efecto o las diferencias entre estos sujetos, así no hagan parte de una situación experimental. De esta manera y sin necesidad de colocar estudiantes en situaciones experimentales éticamente inapropiadas, los resultados que un estudiante obtendría si estuviera en el grupo de tratamiento pueden ser estimados. Esto se logra por medio del emparejamiento de casos con características (covariables) similares sin importar si se trata de casos que reciben o no el tratamiento. Para esto se usan puntajes de propensión o probabilidades de pertenecer a un grupo dadas ciertas características de los sujetos. Este procedimiento es esencial en el PSM y se explicará más adelante.

El PSM ha sido usado con éxito para responder a preguntas sobre el efecto de programas sociales (i.e., Hsieh y Hamilton, 2014), y puede ser empleado también en el campo educativo. Con esta técnica se podrían resolver preguntas de investigación interesantes a partir de datos provenientes de bases nacionales educativas o con datos provenientes de estudios observacionales: ¿cómo sería el desempeño de estudiantes de instituciones de baja calidad educativa si se les colocara en instituciones de alta calidad?, ¿cómo sería el desempeño de estudiantes de instituciones públicas si se les

asignara a instituciones privadas? ¿cómo se vería afectado el logro de los estudiantes de jornada parcial si se les transfiere a jornada completa?

Aplicación de la técnica

Para aplicar la técnica de PSM el primer paso es calcular el PS (puntaje de propensión o "propensity score") por medio de regresión logística en el caso de que la variable independiente sea de tipo bivariado (por ejemplo, pertenecer a una escuela de alta calidad o baja calidad; participar en una escuela de jornada completa o parcial; atender una institución educativa pública o privada). Es decir, se determinará la probabilidad de que los estudiantes pertenezcan a una institución educativa (variable independiente) y no a otra, dadas ciertas características y covariables. El modelo de probabilidad que subyace al PSM es:

$Pr(D \square 1|Xi) \square \pi i \; \square f(\beta Xi)$

Este es un modelo de probabilidad binomial (Pr) de pertenecer a un grupo tratamiento (por ejemplo, colegio de jornada completa) dadas unas características X (las covariables), y que es una función f de los predictores lineales β Xi , donde πi es la probabilidad esperada del estudiante i. Los coeficientes β, estimados mediante máxima verosimilitud, capturan el efecto de las variables X sobre la probabilidad de asistir a un colegio de ciertas características (i.e., jornada completa). Lo que se va a determinar es, ¿cuál sería el logro escolar de los estudiantes de los colegios de jornada completa Y | X, D = 1, si pertenecieran a un colegio sin esta característica (por ejemplo, de media jornada) Y1 | X, D = 0?

La diferencia entre estos logros escolares (el de la jornada completa menos el de la media jornada) es el efecto de pertenecer a un colegio con la característica de interés para el investigador (por ejemplo, la jornada completa). Utilizando los coeficientes β, se construye la distribución de logros escolares para estudiantes de colegios de jornada completa F (Y1 |X, D=1) y media jornada (Y0 |X, D=0), y se utiliza F (Y |X, D=1) para aproximarse a F (Y |X, D=0). En términos de Imbens y Angrist (1994) "debido a que no se observa el mismo individuo bajo las dos condiciones (tratamiento y sin tratamiento) se debe hacer comparaciones entre diferentes individuos y estimar los efectos promedio del tratamiento para poder acercarnos a la estimación del efecto. Es decir, se define el efecto causal en términos de resultados potenciales o contra-factuales" (p. 467).

El cálculo de los puntajes de propensión (que tienen un rango de 0 a 1 por ser una probabilidad) incluirá en la ecuación de regresión todas las covariables que el investigador determine como pertinentes. Las características observables que no se incluyan en el modelo de regresión pueden desmejorar la predicción realizada, por lo que se sugiere que se haga una revisión completa de la literatura en el tema para no dejar variables importantes por fuera del modelo, ya que un número adecuado de estas variables (Hsieh y Hamilton, 2014 incluyen hasta 12) permitirá obtener estimaciones más acertadas de los efectos (Tanaka y Kawakami, 2015).

Aunque para este control de covariables existen técnicas como ANCOVA, no se recomienda su uso si no existen grupos con asignación aleatoria de los cuales obtener la información, ya que ANCOVA no controla por todas las covariables y sus relaciones (Lane y Gibss, 2015).

Una vez obtenidos los puntajes de propensión, estos serán usados para hacer emparejamiento entre las instituciones educativas (una para el grupo tratamiento y otra para el control) que se asemejen en dichos puntajes. Para este emparejamiento, los software estadísticos (R o SPSS) permiten aplicar un procedimiento de emparejamiento exacto a toda la muestra seleccionada, es decir los colegios del grupo tratamiento

tendrán un par o colegio del grupo control con PS idéntico. Otras formas de emparejamiento incluyen: emparejamiento con la puntuación más cercana, con un caliper (número que indica una diferencia máxima entre los PS) o el emparejamiento con el uso de distancias (i.e., Mahalanobis). De esta manera, se podrá distinguir cuáles instituciones tienen influencia en el desempeño de los estudiantes comparadas con un grupo control de instituciones que se les han emparejado estadísticamente pero que no reciben el tratamiento. Es importante tener en cuenta que el procedimiento no se limita a instituciones educativas, sino que se pueden utilizar otras unidades de análisis como los estudiantes.

Después del proceso de emparejamiento de los casos se comprobará que estos no presentan sesgo en el emparejamiento, por ejemplo, graficando un histograma de la distribución de los puntajes de propensión y los casos del grupo tratamiento y control que les corresponden a cada uno de estos puntajes. También se verifica que haya reducción en términos del promedio y de las diferencias estándar (d de Cohen) en los puntajes de propensión y de las covariables (comparación pre y post emparejamiento), para establecer que el emparejamiento fue exitoso. Con la muestra emparejada se puede hacer el cálculo del efecto promedio del tratamiento, el cual se basa en las diferencias entre el promedio de los resultados de los individuos del grupo tratamiento y el promedio de los resultados del grupo control que se le empareja con base a las covariables.

Los casos que no se emparejan se desechan y aquellos que fueron emparejados se emplean en una nueva base de datos para hacer análisis estadísticos de interés para el investigador. Por ejemplo, se puede ingresar en esta nueva base una variable de resultado (o variable dependiente), como desempeño en matemáticas, para establecer si las escuelas de jornada completa (grupo tratamiento) tienen mejor rendimiento comparadas con las escuelas emparejadas del grupo control (escuelas de media jornada). Esto se puede hacer aplicando de nuevo un modelo de regresión (el primero permitió el cálculo de los puntajes de propensión explicado anteriormente). El segundo análisis de regresión incluirá la variable dependiente (los puntajes en matemáticas), la variable independiente jornada (completa o media), junto con todas las covariables que sirvieron para efectos del emparejamiento: tamaño de la escuela (covariable1), porcentaje de estudiantes de minoría (covariable2), precio de la matrícula (covariable3), etc.

De acuerdo con la WWC (What Works Clearinghouse, 2012), en un cuasi-experimento, como aquellos en los que se usan técnicas como PSM, debe existir mínimo dos grupos diferenciados: el grupo que recibe la intervención (por ejemplo, instituciones de jornada completa, de tipo público o de alta calidad) y el grupo control (instituciones de media jornada, o de tipo privado o de baja calidad). Debido a que los grupos pueden diferir, se deben igualar en las características observables lo cual puede ser una fuente de sesgo. En la primera fase del análisis estadístico el criterio de similitud entre el grupo tratamiento y el grupo control consiste en que la diferencia de las características entre los mismos debe ser menor de 0.25 de una desviación estándar, por ejemplo al comparar las covariables (Ho, Inía, King, Stuart, 2007 en WWC, 2012).

Otros estándares que deben cumplir los diseños cuasi-experimentales son: inclusión de una variable dependiente válida y confiable (por ejemplo, una medida estandarizada en la población como los exámenes de último año de Bachillerato); existencia de más de una unidad de análisis en cada grupo (se sugiere un mínimo de 500 unidades en el grupo tratamiento), y garantía de que los efectos se atribuyen a la intervención de acuerdo con la literatura empírica en el área. Debido a las debilidades que en términos de su capacidad de inferencia tienen los estudios cuasi-experimentales, Sadish, Cook y Campbell (2002) recomiendan que en la estimación de efectos se

consideren elementos del diseño como la inclusión de puntuaciones pre-test, el emparejamiento con varios métodos para comparar sus resultados, y hacer uso de múltiples grupos de control.

Ejemplo empírico

Ten Brummelhuis y Vander Lippe (2010, citados por Lane 2015) presentan un ejemplo de la aplicación de la técnica de PSM en el área de RRHH (Recursos Humanos). Los autores utilizan como variable independiente la adopción de horarios flexibles de trabajo (FWA) y como dependiente la productividad y los comportamientos de ayuda a otros por parte de los empleados. Ya que no todo empleado estará satisfecho con este tipo de arreglo flexible se hace necesario identificar las características de quienes obtienen más beneficio con este estilo de trabajo (i.e, trabajadores casados), y así poder emitir una conclusión válida con respecto al efecto de este tipo de programas. Por tanto, se usan las covariables género, edad, educación, tareas del hogar, motivación del empleado, entre otras (tabla 1). Estas son características que se pueden relacionan con el hecho de que una persona asuma o no un horario flexible. Para una muestra de 482 trabajadores, se logra hacer un emparejamiento de 344 casos con PSM, y se encuentran las diferencias antes y después del emparejamiento.

Como se puede observar en la tabla 1, las diferencias entre los grupos de trabajadores con horarios flexibles y no flexibles se disminuyen después del emparejamiento para variables como autonomía y horas laborales. Estas variables parecían tener un efecto mayor en la comparación entre grupos antes del emparejamiento pero sus valores d (diferencia estandarizada) disminuyen después del emparejamiento (i.e., autonomía desciende de 0.67 a 0,08). Así mismo, la edad y el género parecen ganar en importancia en la selección de un horario de trabajo flexible después del emparejamiento ya que sus valores d aumentan (i.e., el efecto de la edad se convierte en positivo e incrementa de -0.3 a 0.04).

Tabla 1. Medias, desviaciones y diferencias medias estandarizadas (d) en las covariantes y puntajes de propensión antes y después del emparejamiento

	Antes del emparejamiento				Después del emparejamiento			
Variable	Media FWA	Medias no FWA	Desviaciones no FWA	d	Media FWA	Medias no FWA	Desviaciones no FWA	d
Puntaje Propensión	0.69	0.32	0.25	1.54	0.69	0.69	0.23	0.04
Mujer	0.46	0.71	0.47	-0.5	0.46	0.37	0.48	0.19
Edad	37.3	41.5	10.5	-0.3	37.3	36.8	12.54	0.04
Autonomía	3.92	3.39	0.74	0.67	3.92	3.85	0.74	0.08
Tareas de casa	9.94	13.45	7.88	-0.4	9.95	10.15	8.09	-0.0
Horas laborales	38.5	29.2	11.21	0.90	38.5	41.80	13.03	-0.3
Presión laboral	3.08	3.17	0.57	-0.1	3.08	3.18	0.64	-0.1

Nota: FWA es horario laboral flexible. Tomado de Ten Brummelhuis & Van Der Lippe, 2010.

Para estimar los efectos del horario laboral flexible, los 344 casos emparejados y sus puntuaciones se usan en un modelo de regresión que predice los comportamientos de ayuda del empleado. En este modelo la varianza total explicada de los comportamientos de ayuda fue de 4.7% que es mayor que la encontrada antes del emparejamiento (con una varianza total explicada del 3%). Por tanto, los autores concluyen a partir de la muestra emparejada que existe un efecto de los horarios flexibles en las conductas de ayuda de los empleados. Un análisis de los coeficientes permite determinar que este efecto se da principalmente para las personas solteras (la

variable tiene un peso B= 0.16, el mayor entre las comunalidades encontradas). Este efecto de la variable estado civil soltero no era detectado antes del emparejamiento (coeficiente Beta= -0.39). (Ten Brummelhuis & Van Der Lippe, 2010).

Software para los cálculos con PSM

Para llevar a cabo PSM se pueden usar varios paquetes estadísticos (R, SPSS, etc.). Randolph, Falbe, Kureethara y Balloun (2014) indican que para hacer un PSM con R, primero se debe instalar la versión gratuita de software en el sitio web de "R Project for Statistical Computing" (https://www.r-project.org/). A continuación se requiere instalar y cargar el paquete MatchIT (el cual se encuentra en el menú "Packages"). El paquete Match IT se descarga solo una vez pero se requiere cargarlo cada vez que se haga un análisis estadístico. Para correr el análisis se requiere una variable de agrupación (VI o tratamiento) y una o más variables de emparejamiento o covariables que permiten igualar a los grupos. Un ejemplo de variable tratamiento puede ser la pertenencia a una escuela pública (por oposición a la escuela privada), a la jornada completa (por oposición a la media jornada), o la asignación a un programa o tratamiento (i.e., un programa de prevención del uso de sustancias, un programa de mejoramiento de las habilidades académicas, etc.) que por lo general son variables dummy que se codifican como 1 o 0.

Los datos que se suben a R pueden venir en cualquier formato pero se recomienda que se guarden como archivos .csv (por ejemplo desde Excell). Este archivo debe incluir casos como filas y los valores de cada variable de interés en las columnas (i.e., las covariables y la variable tratamiento). La variable independiente es la variable de agrupación que indica si un caso pertenece a un tratamiento o a un grupo control y las covariables también deben incluirse en el archivo. El código en R para ingresar la información es el siguiente (tomado de Randolph et al., 2014):

```
mydata <- read.csv ("C:/r/newyork.csv")
attach (mydata)
mydata[1:10,]
```

Este código puede ser copiado reemplazando la información entre paréntesis de la primera línea por la ubicación del archivo a analizar ('C' es el drive donde se almacena la información; 'r'es la carpeta y "newyork" el nombre del archivo, que deben ser reemplazados por los que va a usar el investigador). A continuación se desarrolla el emparejamiento y la evaluación de los resultados con el siguiente código al cual se le deben reemplazar las variables incluidas por las que interesen al investigador:

```
m.out = matchit (stw ~ tot + min + dis, data = mydata, method = "nearest", ratio = 1)
summary (m.out)
Plot(m.out, type = "jitter")
Plot(m.out, type = "hist")
```

La primera línea hace el emparejamiento con el comando Matchit, siendo la variable independiente stw que en el ejemplo original de Randolph et al, 2014 se refiere a tener la condición de ser una escuela del programa Schools to Watch (stw). Las covariables son: tot, min, dis, donde "tot" hace referencia al tamaño de la escuela, "min" se refiere al porcentaje de estudiantes de minoría y "dis" se refiere al porcentaje de estudiantes que reciben almuerzo gratis o de precio reducido. Estas variables deben ser reemplazadas por la variable independiente y covariables que el investigador desee emplear en su análisis.

El comando 'ratio' indica que se emparejará uno a uno, es decir una institución educativa del grupo tratamiento con una del grupo control. Existen varios métodos de

emparejamiento: vecino más cercano (arroja las diferencias promedio más bajas entre los grupos), emparejamiento exacto (empareja cada unidad del grupo tratamiento con una unidad del grupo control que tenga exactamente los mismos valores en cada covariable), sub-clasificación (divide el set de datos en subgrupos cuya distribución de covariables es similar en cada subclase); emparejamiento óptimo (minimiza la distancia promedio entre todos los pares emparejados), y emparejamiento genético (emparejamiento con un algoritmo especial).

Los resultados del emparejamiento son guardados en una variable denominada m.out. Las líneas 2 y 3 del código (que inician con el comando plot) imprimen y hacen gráficos de esta variable. Los resultados del emparejamiento arrojan las medias para el grupo tratamiento y control, la diferencia promedio entre los grupos (antes y después del emparejamiento) en todas las variables y covariables incluidas en el modelo, y la desviación estándar con respecto al grupo control. Estos estadísticos deben mostrar disminución en su cantidad después del emparejamiento, de lo contrario este no es exitoso. En el reporte de resultados el software R también se incluyen las diferencias entre cuartiles entre el grupo control y tratamiento para todas las variables en términos de media, mediana y diferencia máxima entre cuartiles (entre menor sea este último número es mejor el emparejamiento).

Una vez que los datos están emparejados se vuelve a crear una base de datos que solo contiene estos casos emparejados. Esto se hace con el código siguiente (tomado de Randolph, et al., 2014):

```
m.data1 <- match.data(m.out)
write.csv(m.data1, file = "C:/r/newyork_nearest100.csv")
```

La primera línea crea un set de datos en R con los datos emparejados eliminando aquellos que no lo están, la segunda línea convierte el set emparejado en un nuevo archivo de tipo csv. En esta línea se debe reemplazar ‘newyork-nearest100’ por el nombre que se le quiere dar al nuevo archivo. También reemplazar ‘r’ por el archivo donde se quiere almacenar esta base. Con esta nueva base se puede hacer cualquier análisis estadístico que interese al investigador. Por ejemplo, se puede determinar si las escuelas de jornada completa (grupo tratamiento) obtienen mejores resultados que las escuelas de media jornada (grupo control) en pruebas estandarizadas de rendimiento, por medio de una nueva regresión que incluya todas las covariables.

Conclusiones

PSM es una técnica que permite obtener una estimación más ajustada, debido a la inclusión de covariables y a la disminución del sesgo de selección, cuando un efecto no se puede calcular con base a un diseño experimental. Es una técnica que hace uso del cálculo de puntajes de propensión o probabilidades de pertenencia a un grupo para establecer el emparejamiento entre casos de grupo tratamiento y un grupo comparación establecidos por medios estadísticos. Esta técnica no implica la manipulación de variables de forma experimental por parte del investigador lo cual puede ser una ventaja en investigación educativa. El software R permite a los investigadores en educación establecer muestras de casos de tratamiento y control emparejadas con base en puntajes de propensión en particular cuando trabajan con datos como los de bases de resultados de pruebas estandarizadas educativas.

Referencias

Agodini, R. Dinaski, M. (2001). *Are experiments the only option? A look at dropout prevention programs*. Mathematica Policy Research Inc.

Dehejia, R.H., Wahba,S. (1999). Causal effects in non-experimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94 (448)*, 1053-1063.

Heinrich, C., Maffioli, A., y Vasquez, G. (2010). *A primer for applying propensity score matching. Impact Evaluation Guidelines. Techinical Notes N IDB-TN-161*. Interamerican Development Bank.

Holmes, W.M (2014). *Using Propensity Scores in Quasi Experimental Designs*. Los Angeles: Sage.

Hsieh, M., Hamilton, Z.K. (2014). Predicting success in residential substance abuse interventions: New Jersey's pre release incarceration alternatives. *Criminal Justice Policy Review, 1-21*.

Imbens, G.W., y Angrist, J.D. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62(2)*, 467-475.

Lane, F.C., Gibss,S. (2015). Propensity score analysis: a secondary data analysis of work-life Policy and performance outcomes. *Advances in Developing Human Resources, 17*(1), 102-116.

Randolph, J.J., Falbe, K., Kureethara-Manuel, A., & Balloun, J.L. (2014) A step by step guide to Propensity Score Matching. *Practial Assessment, Research and Evaluation, 19* (18), 1-6.

Rosenbaum, Rubin (1984) Reducing bias in observational Studies using sub-classification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Tanaka, S., Kawakami, K. (2015). Methodological issues in observational studies and non-randomized controlled trials in oncology in the era of big data. *Japanese Journal of Clinical Oncology*, 45(4), 323-327.

Ten Brummelhuis, L.L., Van Der Lippe, T. (2010). Effective work-life balance support for various household structures. *Human Resource Management, 49*, 173-193.

Vinha, K. (2006). *A primer on Propensity Score Matching estimators*. Documento CEDE 2006-13. ISNN 1657-7191. Obtenido de: https://economia.uniandes.edu.co/components/com_booklibrary/ebooks/d2006-13.pdf

What Works Clearinghouse (2012). *Procedures and Standards Handbook (versión 2.0)*. US Department of Education. Institute of Education Sciences.

Educación y Ciencia